HYBRID SIGNATURE SCHEME

The present invention relates to methods and apparatus for digitally signing a message.

Digital signatures are used to sign a message generated by a correspondent so that the origin and authenticity of the message may subsequently be verified. In its basic form, a digital signature of a message is generated by signing the message with the originators private key. The message may then be recovered using the originators public key. A number of variants of this basic arrangement have been proposed with different attributes. Digital signature schemes are typically thought to fall into two generic classes, namely digital signatures with appendix and digital signatures with message recovery.

Digital signatures with appendix are categorized by the fact that the message signed is required as input to the verification algorithm. Although very popular (the DSS and ECDSA are examples of this mechanism) they may not provide as much bandwidth efficiency as other methods.

Sub a1

~~Digital signatures with message recovery are categorized by the fact that the message is not required as input to the verification algorithm. One problem with message recovery schemes is to defeat existential forgery attacks by defining a suitable redundancy function which will distinguish messages legitimately signed from signatures of random bit strings.~~

In many practical applications the data to be signed carries a certain amount of inherent redundancy. For example, four bytes of data might be reserved for the date but, in practice, 3 bytes suffice and so there are 8 bits of redundancy from this field. In order to ensure security it is necessary to provide a predetermined degree of redundancy within the message and accordingly the bandwidth efficiency is reduced.

To increase the bandwidth efficiency it is known to split the message in to two components, namely a hidden and a visible component. The hidden component is recovered during the verification process and the visible portion is used as an input to the recovery process. The hidden component must have sufficient redundancy to withstand an existential forgery attack and additional bits must be added to the message if it does not inherently possess this. In one of the proposed standards to implement such a scheme,

ISO 9796 Part 2, the hidden component is utilised to generate a signature component c of the form $DES_R[H//SHA1(V)//I_A]$ where

H is the hidden component,

V is the visible component

I_A is an identifier of the signer

SHA1(V) is a cryptographic hash of the visible component, and

DES_R is an encryption of the bit string.

This scheme however has the disadvantage that c is at least the number of bits in SHA1 (V) bits longer, and, as it is included in the signature, the required bandwidth efficiency may not be achieved. Moreover, the scheme requires invocation of two hash operations as the value c is subsequently hashed for inclusion in the signature component. This computational complexity may make it unsuitable for certain applications.

It is therefore an object of the present invention to provide a signature scheme in which the above disadvantages are obviated or mitigated.

In general terms, one aspect of the present invention provides a signature scheme in which a message is divided in to a first portion which is hidden and is recovered during verification, and a second portion which is visible and is required as input to the verification algorithm. A first signature component is generated by encrypting the first portion alone. An intermediate component is formed by combining the first component and the visible portion and cryptographically hashing them. A second signature component is then formed using the intermediate component and the signature comprises the first and second components with the visible portion.

The generation of the first component from the first portion alone reduces the necessary bandwidth and simplifies the computation. The relative sizes of the first and second portions are determined by the application itself. In this manner, the redundancy function can be application dependent as opposed to a global primitive.

Recovery of the message can be completed using the signature and the public key of the sender.

According to a further aspect of the invention there is provided a verification of a signature of a message that has been subdivided into a hidden and visible portion. The verification combines a first component derived only from the hidden portion of the

message with the visible portion and produces a hash of the combination. The computed hash is used together with publicly available information to generate a bit string corresponding to the hidden portion. If the required redundancy is present the signature is accepted and the message reconstructed from the recovered bit string and the visible portion.

Embodiments of the invention will now be described by way of example only with reference to the accompanying drawings in which: -

Figure 1 is a schematic representation of a data communication system,

Figure 2 is a flow chart showing the signature generation,

Figure 3 is a flow chart showing the verification of the signature of figure 2, and

Figure 4 is a flow chart showing a further embodiment of signature generation.

Referring to Figure 1, a data communication system includes a pair of correspondents 10, 12 exchanging a message M over a communication channel 14. Each of the correspondents 10, 12 includes a cryptographic unit 16, 18 respectively and a terminal 20, 22 to generate and receive the message M. Each of the cryptographic units 16, 18 implements a public key encryption scheme that enables it to generate a session key, to encipher or decipher a message using the session key or sign a message using a private key which can then be recovered using a corresponding public key. The general implementation of such schemes and their operating principles are well known. The encryption scheme may be loaded in to the encryption unit from a data carrier coded to implement the protocol under the direction of a general purpose computer or may be implemented on a chipset as preprogrammed instructions.

In the preferred embodiment described below, the encryption scheme is based on the intractability of the discrete log problem in finite groups and is implemented in an algebraic system defined on the points of an elliptic curve over a finite field, typically referred to as elliptic curve crypto systems. However, the signature scheme proposed may be applied to any ElGamal signature over any finite group.

The domain parameters of such an elliptic curve crypto system are a curve of the form $y^2 = x^3 + dx + c$ and a seed point P. One of the correspondents has a private key a,

$0<a<n$ where n is the order of the point P and a corresponding public key $Q_A = aP$. The public key may be held in a certifying authority 24 shown in communication with the correspondents 10, 12 by ghosted lines.

The messages M generated by the correspondents 10, 12 are subdivided into two bit strings H and V (i.e. M=H//V) where H is a bit string which is hidden and recovered during the verification process and V is a bit string which is also signed but is required as input to the verification process.

The signature generation algorithm is set out in the flow chart of figure 2. Initially the bit string H is examined to determine if it contains redundancy above a predetermined limit sufficient to prevent an existential forgery attack. If the examination determines that the original data forming the message M contains enough redundancy then H may simply be a subset of that data. If the predetermined redundancy is not found then H may be modified to contain artificially added redundancy such as additional bytes of O's.

By way of example, suppose 80 bits of redundancy is determined to be the predetermined lower limit for security reasons. If the bit string H contains no inherent redundancy then it would be necessary to add up to 10 bytes of 0's. To permit recovery of the message an indicator would be included, conveniently as a leading byte in either H or V, which tells the number of bytes of 0's added. Since the value is 0 to 10, 4 bits of the byte suffice as an indicator so the bit string contains an additional 4 bits of redundancy. If t is the number of redundancy bytes that can be added, then the data must inherently contain at least 80-8t bits of redundancy.

To sign the message M = H // V the correspondent 10 generates a random integer k, $o<k<n$ in the cryptographic unit 14. Using k correspondent 10 then computes a value of a random point R = kP.

A value c is then computed from the bit string H only such that $c = SKE_R (H)$. SKE_R refers to a symmetric-key algorithm under control of a key derived from the

random point R. This could be derived by applying a function, such as a hash function, to R, truncating R, or using only one of the coordinates, e.g. the x coordinate as the key. If H is smaller than the key derived from R, then one possible SKE is simply to XOR H with a truncation of bits from the key derived from R. This effectively is a one-time pad. If H is larger than the key it is possible to use a DES based algorithm or simply to XOR repeatedly the key with H.

Using the bit string V, an intermediate component c' is computed such that c'= SHA1 (c//V) where SHA1 is a cryptographically secure hash algorithm. If preferred, additional information such as a certificate or identifying information of correspondent 10 may be incorporated in to the hashed value c'.

It will be noted that the signature component c is the same length as the hidden portion H as it is a bit wise encryption of that portion and that the intermediate component c' is obtained with a single hash operation.

A signature component s is then computed from the values available to the correspondent 10 using any of the known ElGamal equations. A convenient equation is the Schnorr signature algorithm where s = c'a+k (mod n). A signature is then formed from the components (s,c,V) and forwarded to the correspondent 12.

Verification of the signature by correspondent 12 is performed by the application of the corresponding algorithm, as shown in figure 3 for the Schnorr signature. The correspondent 12 initially obtains an authentic copy of the public key Q_A of the correspondent 10 from the certifying authority 24. The correspondent 12 then computes a value c"= SHA1 (c//V) and derives from the information available in the signature, i.e. s,c,V and the system domain parameters, the values

$$X = sP$$

$$Y = c'' Q_A$$

$$Z = X - Y$$

A bit string H' is then recovered by applying to the received signature component c the symmetric-key algorithm under control of a key derived from the point Z such that H' = SKE_Z (c). The bit string H' is then examined to determine if it has the required redundancy and if so the correspondent 12 accepts the signature of M and reconstitutes the message as H'//V.

Because the message M is subdivided, it is only necessary for the one portion, H, to contain the requisite redundancy. The other portion V, which is sent in the clear, may have the data structure of the original data and thereby improve the bandwidth efficiency.

Another feature of this scheme which is of practical and commercial interest is that the information encoded in c is only available to those individuals who have the public key Q_A of correspondent 10. The data contained in V is available to all. There may be some information which correspondent 10 wants to hide from those not privy to Q_A in which case the sender, i.e. correspondent 10 puts this information into the bit string H.

For example, in one particular application where the signature is used to authenticate postage applied to mail, a mailer may not want a receiver to know how many mail pieces he has sent. The post office (which verifies postage and therefore needs this information) has the public key of the mailer, and can recover this information on verification but the receiver cannot if he does not have the mailers public key.

Of course, if the public key Q_A of the sender is contained in the indicium then this is also available to the receiver. Alternatively, the senders public key may be contained in a certificate that can only be recovered if the receiver has the certifying authority's public key. If this is not generally available then the contents of H will be hidden from the receiver.

As indicated above, alternative forms of signing equations may be used. In a further embodiment shown in the flow chart of figure 4, a signing equation similar to the ECDSA standard is used. Normally in such an arrangement:-

- $R = kP$
- $c = DES_R(M)$
- $r' = SHA1(c)$
- $s = k^{-1}\{SHA1(c//ID_A) + a\,r'\} \bmod n$ where ID_A is an identifier of the sender.
- the signature is (s,c).

When used with a hybrid scheme described above the scheme is modified such that

- $R = kP$
- $c = DES_R(H)$
- $r' = SHA1(c)$
- $s = k^{-1}\{SHA1(c//V) + a\,r'\}$ modn.
- the signature is (s, c,V)

Again therefore because only a portion H of the message is used to generate the first component c, only that portion requires a specified redundancy. In the balance of the message a reduced redundancy may be utilised to maintain bandwidth efficiency.

The verification for the modified scheme will change accordingly to accommodate the partial message recovery and necessary redundancy.